Query Performance Prediction: From Fundamentals to Advanced Techniques

Negar Arabzadeh[1(✉)], Chuan Meng[2], Mohammad Aliannejadi[2], and Ebrahim Bagheri[3]

[1] University of Waterloo, Waterloo, Canada
narabzad@uwaterloo.ca

[2] University of Amsterdam, Amsterdam, The Netherlands
{c.meng,m.aliannejadi}@uva.nl

[3] Toronto Metropolitan University, Toronto, Canada
Bagheri@torontomu.ca

Abstract. Query performance prediction (QPP) is a core task in information retrieval (IR) that aims at predicting the retrieval quality for a given query without relevance judgments. QPP has been investigated for decades and has witnessed a surge in research activity in recent years; QPP has been shown to benefit various aspects, e.g., improving retrieval effectiveness by selecting the most effective ranking function per query [5,7]. Despite its importance, there is no recent tutorial to provide a comprehensive overview of QPP techniques in the era of pre-trained/large language models or in the scenario of emerging conversational search (CS); In this tutorial, we have three main objectives. First, we aim to disseminate the latest advancements in QPP to the IR community. Second, we go beyond investigating QPP in ad-hoc search and cover QPP for CS. Third, the tutorial offers a unique opportunity to bridge the gap between theory and practice; we aim to equip participants with the essential skills and insights needed to navigate the evolving landscape of QPP, ultimately benefiting both researchers and practitioners in the field of IR and encouraging them to work around the future avenues on QPP.

1 Motivation and Objectives

Query performance prediction (QPP) is a core task in information retrieval (IR). The task of QPP aims to predict the effectiveness of retrieved results for a given query in the absence of relevance judgments [14,18,24,33]. The importance of the QPP task cannot be overstated because of numerous applications of QPP. QPP serves as a guiding compass at query time. QPP can choose the most effective ranking function per query and select the best variant from multiple query reformulations [51], delivering more relevant and tailored results to users [11,15]. QPP can be instrumental in optimizing the retrieval efficiency by choosing the appropriate number of documents to process in a multi-stage retrieval setting [19]. Moreover, in the scenario of emerging conversational search (CS) [25,37,39], QPP can predict the retrieval quality for each user query in a conversation and

N. Goharian et al. (Eds.): ECIR 2024, LNCS 14612, pp. 381–388, 2024.
https://doi.org/10.1007/978-3-031-56069-9_51

assist an intelligent system in determining when to take the initiative (e.g., ask a clarifying question) [1] to get more information from a user [6,43]. Apart from the application at query time, a recent study [29] has shown the application of QPP in creating a collection; this study uses QPP to predict adaptive pool depth for each query so as to reduce relevance judgment costs.

Despite the importance of QPP to the IR community, to the best of our knowledge, there has been only one tutorial [42] about QPP at ICTIR 2020. We believe that it is the right time to revisit the idea of having a QPP tutorial for the following three reasons: (i) with the rapid development of advanced neural-based techniques in recent years, QPP has witnessed a surge in proposing neural-based QPP methods [4,16,23,35,36,47,51] or predicting the performance of neural-based retrievers [3,22,26,27,47]; (ii) conversational search (CS) [30,38,52] has been recognized as an emerging research area in IR, and research [25,37,39] into QPP for CS have been conducted very recently.

This tutorial has three main aims to provide a platform to disseminate the latest advancements in QPP to the IR community. Moreover, we go beyond investigating QPP in ad-hoc search and delve into studying QPP for CS, which has remained relatively under-explored. Last but not least, this tutorial offers a unique opportunity to bridge the gap between theory and practice. We will provide practical insights and hands-on experience to empower researchers, practitioners, and enthusiasts to effectively apply QPP techniques in their work. In summary, our main objectives in this tutorial are as follows:

– **Highlighting the importance of QPP:** we aim to underscore the pivotal role that QPP plays in enhancing IR, by illuminating various applications of QPP.
– **Providing a comprehensive overview of latest QPP methods:** our tutorial endeavors to provide participants with a thorough and up-to-date overview of state-of-the-art QPP techniques. This comprehensive survey spans both lexical-based methods and the latest neural-based QPP approaches, ensuring that attendees gain insights into the diverse landscape of QPP methodologies.
– **Exploring the impact of cutting-edge neural-based techniques advancements on QPP:** we shed light on the latest developments within the QPP task, specifically, how predicting the performance for neural-based retrievers differs from conventional sparse retrievers. We also cover the work done exploring the potential of leveraging potential of leveraging large language models (LLMs) to enhance QPP methods. Our goal is to foster a forward-looking perspective, inspiring participants to contribute to the ongoing evolution of QPP research in the IR community.
– **Exploring QPP for CS:** we first summarize the findings and insights from very recent studies into QPP for CS, and then show the drawbacks of current QPP methods in the scenario of CS, and discuss open avenues for future research.
– **Facilitating hands-on experience with supporting materials:** to enhance the learning experience, we are committed to equipping participants with supporting materials that enable practical application.

2 Format and Schedule

This tutorial will span a half-day, totaling 3 h plus breaks, and will be presented in person. We propose the following program:

Introduction to QPP [25 min]. This opening section provides (i) a foundational understanding of the task of QPP in the context of ad-hoc search and CS, (ii) applications of QPP (why QPP is vital) in these domains, (iii) different categorizations of QPP methods, including pre-retrieval and post-retrieval QPP methods, and (iv) various evaluation methods [28] relevant to QPP task.

Pre-retrieval QPP Methods [30 min]. This section focuses on strategies for conducting QPP before the retrieval process. We include (i) statistical-based pre-retrieval QPP methods [32–34,53], and (ii) recently introduced neural-based ones [8,9,44,45,50] We highlight the drawbacks and advantages of both categories.

Post-retrieval QPP Methods [55 min]. Since post-retrieval QPP methods have attracted more attention compared to pre-retrieval QPP methods, we will cover both lexical-based QPP methods and neural-based methods, separately. We will explore the principles and techniques that underlie lexical-based QPP, gaining insights into how lexical features and linguistic analysis can be harnessed to assess and enhance QPP after retrieval, covering state-of-the-art retrieval score-based QPP methods [10,20,40,46,48,54]. In addition, We dive into the cutting-edge domain of neural-based post-retrieval QPP methods. Participants will acquire a profound understanding of how neural networks and advanced machine learning techniques are employed to forecast query performance following the retrieval process. Our exploration covers QPP methods based on basic neural networks [2,21,51], along with methods based on pre-trained language models [4,16,23,31,36].

Impact of Retriever Types on QPP Effectiveness [25 min]. Many studies in QPP have primarily focused on predicting the performance of traditional high-dimensional sparse retrievers like BM25. However, recent research has expanded the scope by investigating how predicting the performance of other retriever types, such as dense or learned sparse retrievers, can differ from sparse retrievers [3,22,26,27,47]. In this section, we first demonstrate and discuss how different retriever types impact the effectiveness of QPP methods, paving the way for a deeper understanding of this evolving landscape; and we pose research questions that hold the potential to shape the future of QPP in this domain, encouraging exploration in this field.

QPP and LLMs [15 min]. LLMs have already been successfully used in many IR tasks, such as re-ranking [41] and query expansion [49]. However, QPP using LLMs has been little studied. In this section, we will explore and discuss the potential ways of harnessing LLMs to enhance QPP effectiveness.

QPP for CS [15 min]. This section delves into QPP for CS. We aim to (i) summarize the findings and insights from very recent studies [25,37,39] into

QPP for CS, (ii) show the drawbacks of current QPP methods in the scenario of CS, and (iii) discuss open avenues for future research.

Limitations and Future Work [10 min]. In the final segment, we will (i) engage in a thoughtful discussion about the major theoretical and conceptual limitations that currently exist in QPP research and practice, and (ii) present exciting avenues for future work, inspiring participants to contribute to the ongoing evolution of QPP in IR. Last but not least, throughout the whole tutorial, we hold interactive *hands-on experience sessions* in which, participants will receive access to interactive Google Colab Notebooks, built on top of our recent comprehensive QPP repository,[1] allowing them to apply their knowledge in practice. Attendees will have the opportunity to implement QPP methods in real-world scenarios, reinforcing their understanding and refining their practical skills, providing a more tangible experience of prediction outcomes and enabling participants to compare them from computational perspectives.

3 Intended Audience

Our tutorial is thoughtfully designed to accommodate a diverse audience, catering to individuals with varying levels of familiarity with IR and related subjects. Our intended audience for this tutorial consists of two main groups: (i) professionals who already possess a solid understanding of fundamental IR techniques; our tutorial serves as a valuable extension of their knowledge by providing deeper insights and practical knowledge; and (ii) people who may be relatively new to these advanced topics; we offer them a friendly entry point by starting with the basics of problem formulation; while we assume that our participants have a foundational understanding of topics typically covered in an undergraduate IR course, our tutorial will provide the necessary details and explanations to ensure that all participants can comfortably access and comprehend the content.

4 Presenters

Negar Arabzadeh is a Ph.D. student at the University of Waterloo. Her research is aligned with ad-hoc search and CS in IR. Negar's Master's thesis was focused on neural-based pre-retrieval QPP. She has published relevant papers in SIGIR, CIKM, ECIR, and IP&M. Negar has previously conducted tutorials in SIGIR 2022, ECIR 2023, and WSDM 2023 [12,13,17].

Chuan Meng is a Ph.D. student at IRLab, University of Amsterdam, supervised by Maarten de Rijke and Mohammad Aliannejadi. His main research topic is CS and QPP. Chuan has published papers in prestigious proceedings such as SIGIR, EMNLP, CIKM, and AAAI. Moreover, he serves as a committee member for conferences including ACL, WWW, EMNLP, WSDM, COLING, SIGKDD, AAAI, ECIR, and a reviewer for journals including TOIS and IP&M.

1 https://github.com/ChuanMeng/QPP4CS.

Mohammad Aliannejadi is an Assistant Professor at IRLab, University of Amsterdam. His research interests include conversational information access, recommender systems, and QPP. Mohammad has co-organized various evaluation campaigns such as TREC CAsT, TREC iKAT, ConvAI3, and IGLU. Moreover, Mohammad has held multiple tutorials and lectures on CS, such as CHIIR, SIKS, and ASIRF.

Ebrahim Bagheri is a Professor and the Director for the Laboratory for Systems, Software, and Semantics (LS^3) at Toronto Metropolitan University. He holds a Canada Research Chair (Tier II) in social information retrieval and an NSERC industrial research chair in social media analytics. He currently leads the NSERC program on responsible AI (http://responsible-ai.ca). He is an associate editor for ACM transactions on intelligent systems and technology (TIST) and Wiley's computational intelligence.

References

1. Aliannejadi, M., Kiseleva, J., Chuklin, A., Dalton, J., Burtsev, M.: Building and evaluating open-domain dialogue corpora with clarifying questions. In: EMNLP (2021)
2. Arabzadeh, N., Bigdeli, A., Zihayat, M., Bagheri, E.: Query performance prediction through retrieval coherency. In: Hiemstra, D., Moens, M.-F., Mothe, J., Perego, R., Potthast, M., Sebastiani, F. (eds.) ECIR 2021, Part II. LNCS, vol. 12657, pp. 193–200. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72240-1_15
3. Arabzadeh, N., Hamidi Rad, R., Khodabakhsh, M., Bagheri, E.: Noisy perturbations for estimating query difficulty in dense retrievers. In: CIKM (2023)
4. Arabzadeh, N., Khodabakhsh, M., Bagheri, E.: BERT-QPP: contextualized pre-trained transformers for query performance prediction. In: CIKM (2021)
5. Arabzadeh, N., Mitra, B., Bagheri, E.: MS MARCO chameleons: challenging the MS MARCO leaderboard with extremely obstinate queries. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 4426–4435 (2021)
6. Arabzadeh, N., Seifikar, M., Clarke, C.L.: Unsupervised question clarity prediction through retrieved item coherency. In: CIKM, pp. 3811–3816 (2022)
7. Arabzadeh, N., Yan, X., Clarke, C.L.: Predicting efficiency/effectiveness trade-offs for dense vs. sparse retrieval strategy selection. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 2862–2866 (2021)
8. Arabzadeh, N., Zarrinkalam, F., Jovanovic, J., Al-Obeidat, F., Bagheri, E.: Neural embedding-based specificity metrics for pre-retrieval query performance prediction. IP&M **57**(4), 102248 (2020)
9. Arabzadeh, N., Zarrinkalam, F., Jovanovic, J., Bagheri, E.: Neural embedding-based metrics for pre-retrieval query performance prediction. In: Jose, J.M., et al. (eds.) ECIR 2020, Part II. LNCS, vol. 12036, pp. 78–85. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45442-5_10
10. Arabzadeh, N., Zarrinkalam, F., Jovanovic, J., Bagheri, E.: Geometric estimation of specificity within embedding spaces. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2109–2112 (2019)

11. Bellogín, A., Castells, P.: Predicting neighbor goodness in collaborative filtering. In: Andreasen, T., Yager, R.R., Bulskov, H., Christiansen, H., Larsen, H.L. (eds.) FQAS 2009. LNCS (LNAI), vol. 5822, pp. 605–616. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04957-6_52
12. Bigdeli, A., Arabzadeh, N., SeyedSalehi, S., Zihayat, M., Bagheri, E.: Gender fairness in information retrieval systems. In: SIGIR (2022)
13. Bigdeli, A., Arabzadeh, N., Seyedsalehi, S., Zihayat, M., Bagheri, E.: Understanding and mitigating gender bias in information retrieval systems. In: Kamps, J., et al. (eds.) ECIR 2023. LNCS, vol. 13982, pp. 315–323. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28241-6_32
14. Carmel, D., Yom-Tov, E.: Estimating the query difficulty for information retrieval. Synthesis Lectures on Information Concepts, Retrieval, and Services (2010)
15. Carmel, D., Yom-Tov, E., Roitman, H.: Enhancing digital libraries using missing content analysis. In: Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 1–10 (2008)
16. Chen, X., He, B., Sun, L.: Groupwise query performance prediction with BERT. In: Hagen, M., et al. (eds.) ECIR 2022. LNCS, vol. 13186, pp. 64–74. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99739-7_8
17. Clarke, C.L., Diaz, F., Arabzadeh, N.: Preference-based offline evaluation. In: WSDM, pp. 1248–1251 (2023)
18. Cronen-Townsend, S., Zhou, Y., Croft, W.B.: Predicting query performance. In: SIGIR, pp. 299–306 (2002)
19. Culpepper, J.S., Clarke, C.L., Lin, J.: Dynamic cutoff prediction in multi-stage retrieval systems. In: Proceedings of the 21st Australasian Document Computing Symposium, pp. 17–24 (2016)
20. Cummins, R., Jose, J., O'Riordan, C.: Improved query performance prediction using standard deviation. In: SIGIR (2011)
21. Datta, S., Ganguly, D., Greene, D., Mitra, M.: Deep-QPP: a pairwise interaction-based deep learning model for supervised query performance prediction. In: WSDM, pp. 201–209 (2022)
22. Datta, S., Ganguly, D., Mitra, M., Greene, D.: A relative information gain-based query performance prediction framework with generated query variants. TOIS **41**(2), 1–31 (2022)
23. Datta, S., MacAvaney, S., Ganguly, D., Greene, D.: A 'pointwise-query, listwise-document' based query performance prediction approach. In: SIGIR, pp. 2148–2153 (2022)
24. Faggioli, G., Ferro, N., Mothe, J., Raiber, F.: QPP++ 2023: query-performance prediction and its evaluation in new tasks. In: Kamps, J., et al. (eds.) ECIR 2023. LNCS, vol. 13982, pp. 388–391. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28241-6_42
25. Faggioli, G., Ferro, N., Muntean, C.I., Perego, R., Tonellotto, N.: A geometric framework for query performance prediction in conversational search. In: SIGIR, pp. 1355–1365 (2023)
26. Faggioli, G., et al.: Towards query performance prediction for neural information retrieval: challenges and opportunities. In: ICTIR, pp. 51–63 (2023)
27. Faggioli, G., Formal, T., Marchesin, S., Clinchant, S., Ferro, N., Piwowarski, B.: Query performance prediction for neural IR: Are we there yet? In: Kamps, J., et al. (eds.) ECIR 2023. LNCS, vol. 13980, pp. 232–248. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-28244-7_15

28. Faggioli, G., Zendel, O., Culpepper, J.S., Ferro, N., Scholer, F.: sMARE: a new paradigm to evaluate and understand query performance prediction methods. Inf. Retr. J. **25**(2), 94–122 (2022)
29. Ganguly, D., Yilmaz, E.: Query-specific variable depth pooling via query performance prediction. In: SIGIR, pp. 2303–2307 (2023)
30. Gao, J., Xiong, C., Bennett, P., Craswell, N.: Neural approaches to conversational information retrieval. arXiv preprint arXiv:2201.05176 (2022)
31. Hashemi, H., Zamani, H., Croft, W.B.: Performance prediction for non-factoid question answering. In: SIGIR, pp. 55–58 (2019)
32. Hauff, C.: Predicting the effectiveness of queries and retrieval systems. In: SIGIR Forum, vol. 44, p. 88 (2010)
33. Hauff, C., Hiemstra, D., de Jong, F.: A survey of pre-retrieval query performance predictors. In: CIKM, pp. 1419–1420 (2008)
34. He, B., Ounis, I.: Inferring query performance using pre-retrieval predictors. In: International Symposium on String Processing and Information Retrieval (2004)
35. Khodabakhsh, M., Bagheri, E.: Semantics-enabled query performance prediction for ad hoc table retrieval. Inf. Process. Manage. **58**(1), 102399 (2021)
36. Khodabakhsh, M., Bagheri, E.: Learning to rank and predict: multi-task learning for ad hoc retrieval and query performance prediction. Inf. Sci. **639**, 119015 (2023)
37. Meng, C., Aliannejadi, M., de Rijke, M.: Performance prediction for conversational search using perplexities of query rewrites. In: QPP++2023, pp. 25–28 (2023)
38. Meng, C., Aliannejadi, M., de Rijke, M.: System initiative prediction for multi-turn conversational information seeking. In: CIKM, pp. 1807–1817 (2023)
39. Meng, C., Arabzadeh, N., Aliannejadi, M., de Rijke, M.: Query performance prediction: From ad-hoc to conversational search. In: SIGIR, pp. 2583–2593 (2023)
40. Pérez-Iglesias, J., Araujo, L.: Standard deviation as a query hardness estimator. In: Chavez, E., Lonardi, S. (eds.) SPIRE 2010. LNCS, vol. 6393, pp. 207–212. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16321-0_21
41. Pradeep, R., Sharifymoghaddam, S., Lin, J.: RankVicuna: zero-shot listwise document reranking with open-source large language models. arXiv preprint arXiv:2309.15088 (2023)
42. Roitman, H.: ICTIR tutorial: modern query performance prediction: theory and practice. In: ICTIR, pp. 195–196 (2020)
43. Roitman, H., Erera, S., Feigenblat, G.: A study of query performance prediction for answer quality determination. In: SIGIR, pp. 43–46 (2019)
44. Roy, D., Ganguly, D., Mitra, M., Jones, G.J.: Estimating gaussian mixture models in the local neighbourhood of embedded word vectors for query performance prediction. IP&M **56**(3), 1026–1045 (2019)
45. Salamat, S., Arabzadeh, N., Seyedsalehi, S., Bigdeli, A., Zihayat, M., Bagheri, E.: Neural disentanglement of query difficulty and semantics. In: CIKM, pp. 4264–4268 (2023)
46. Shtok, A., Kurland, O., Carmel, D., Raiber, F., Markovits, G.: Predicting query performance by query-drift estimation. TOIS **30**, 1–35 (2012)
47. Singh, A., Ganguly, D., Datta, S., McDonald, C.: Unsupervised query performance prediction for neural models with pairwise rank preferences. In: SIGIR, pp. 2486–2490 (2023)
48. Tao, Y., Wu, S.: Query performance prediction by considering score magnitude and variance together. In: CIKM (2014)
49. Wang, L., Yang, N., Wei, F.: Query2doc: query expansion with large language models. arXiv preprint arXiv:2303.07678 (2023)

50. Zamani, H., Bendersky, M.: Multivariate representation learning for information retrieval. arXiv preprint arXiv:2304.14522 (2023)
51. Zamani, H., Croft, W.B., Culpepper, J.S.: Neural query performance prediction using weak supervision from multiple signals. In: SIGIR (2018)
52. Zamani, H., Trippas, J.R., Dalton, J., Radlinski, F.: Conversational information seeking. arXiv preprint arXiv:2201.08808 (2022)
53. Zhao, Y., Scholer, F., Tsegay, Y.: Effective pre-retrieval query performance prediction using similarity and variability evidence. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) ECIR 2008. LNCS, vol. 4956, pp. 52–64. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78646-7_8
54. Zhou, Y., Croft, W.B.: Query performance prediction in web search environments. In: SIGIR, pp. 543–550 (2007)